W412–W416 *Nucleic Acids Research, 2013, Vol. 41, Web Server issue* *Published online 22 April 2013*
doi:10.1093/nar/gkt299

Adepth: new representation and its implications for atomic depths of macromolecules

Shu-wen W. Chen[1] and Jean-Luc Pellequer[2,*]

[1]13 Avenue de la Mayre, F-30200 Bagnols sur Cèze, France and [2]CEA, iBEB, Service de Biochimie et Toxicologie Nucléaire, F-30207 Bagnols sur Cèze, France

Received January 20, 2013; Revised March 15, 2013; Accepted March 31, 2013

ABSTRACT

We applied the signed distance function (SDF) for representing the depths of atoms in a macromolecule. The calculations of SDF values were performed on grid points in a rectangular box that accommodates the macromolecule. The depth for an atom inside the molecule was then obtained as a result of tri-linear interpolation of SDF values at the nearest grid points surrounding the atom. For testing the performance of present program Adepth, we have constructed an artificial molecule whose atomic depths are known as the gold standard for accuracy assessments. On average, our results showed that Adepth reached an accuracy of 1.6% at 0.5 Å of grid spacing, whereas the current reference server DEPTH reached 7.5%. The Adepth program provides both depth and height representations; it is capable of computing iso-surfaces for atomic depths and presenting graphical view of macromolecular shape at some distance away from the surface. Web interface is available at http://biodev.cea.fr/adepth.

INTRODUCTION

The atomic depth has been found extremely useful in studying hydrogen/deuterium exchange rates (1), the characterization of protein domain size and identification of potential targets for phosphorylation (2), quantifications of driving force for a burial residue (3), protein fold recognition (4) and predictions of small-molecule binding cavities in proteins (5,6). As a closely related quantity, residue depth was used to study mutational impacts on protein stability (7) and the relationship of depth measure with protein sequences (8,9).

In the late 70', macromolecular shape was described by a map of solvent exclusion, a method that could claim the first computation of atomic depths (10). Since then several methods have been devised to compute atomic depths in proteins. The first one used a pre-equilibrated solvent-bath in which the macromolecule of interest is immerged. Water molecules that bump with macromolecular atoms are removed from the bath, and the depth is assigned as a value of the distance between any atom and the nearest water molecule on the surface (7). Several other methods require computations of ASA to determine which atoms are buried beforehand (11). The depth of a buried atom is then measured as the distance between the atom and its closest solvent-accessible atom (2) or the closest vertex of ASA surface (8,9). Alternatively, one may attribute the depth index to each atom in the macromolecule by measuring the solvent-exposed volume of a virtual sphere with a varying radius located at the atomic centre (12).

In this work, we present an unsupervised program, Adepth, which computes the atomic depth without requiring previous knowledge of ASA or preparations of a solvent bath. Another advantage of Adepth relies in its integrity on identification of surface atoms and quantification of atomic depths in one run. Validation of the program was performed on a well-designed molecular structure with known depth information (Figure 1). Applications of Adepth includes definition of protein skins with a user-defined thickness, calculations of atom–atom overlaps in macromolecular assemblies (14) and a graphical visualization of depth using iso-surfaces in VMD (13). Another useful application of Adepth concerns the assessment of tip-convolution effects in atomic force microscopy (AFM) imaging (15). In the range of grid spacing 0.5–1 Å, Adepth has been shown to be more accurate than available servers that were tested in the present study.

IMPLEMENTATION

In the earlier work (16), the signed distance function (SDF) field of a protein allowed us to identify solvent accessible atoms and determine how deep a residue is inside the protein. In summary, we placed the protein in a rectangular grid of which width, length and height were determined by the size of the protein. The grid was

*To whom correspondence should be addressed. Tel: +33 466 79 1943; Fax: +33 466 79 1905; Email: jlpellequer@cea.fr

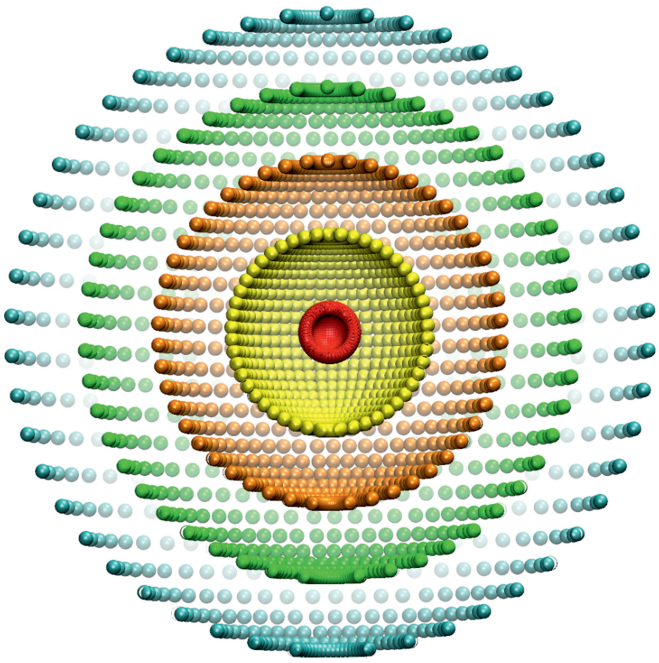

Figure 1. Representation of the 3D structure of the standard molecule drawn by the vdw method in VMD (13). Each atom was scaled down to 20% of its original size. The Cartesian co-ordinates of atoms in the molecule can be expressed as (Rcosφsinθ, Rsinφsinθ, Rcosθ), where R is the radius of sphere, $0 \le \theta \le \pi$ and $0 \le \varphi < 2\pi$; totally there are 25 sampling points along the θ coordinate and 50 for the φ coordinate. Atomic depth is colour-coded as follows: cyan = 0 Å, green = 3 Å, orange = 6 Å, yellow = 9 Å and red = 12 Å.

uniformly discretized in three dimensions at user-defined grid spacing (Δx, Δy and Δz). For each atom of the protein, the atomic van der Waals (vdw) radius was given based on the atomic type defined in the Charmm22 force field (17), as shown in Table 1. By default, the radius of solvent or probe was set to 1.5 Å. For representing the solvent-accessible surface of the molecule, an extended radius, the sum of vdw and solvent radii, was assigned to each heavy atom. Following level set methods (18), we first computed the SDF values at interfacial grid points (level 0) using information of where are limits of atomic and solvent force fields. For a grid point (i j k) at next level, the SDF value, $D^S(i\ j\ k)$, was obtained by solving the quadratic equation (19)

$$\frac{\left[D^S(ijk) - D_1\right]^2}{\Delta x^2} + \frac{\left[D^S(ijk) - D_2\right]^2}{\Delta y^2} + \frac{\left[D^S(ijk) - D_3\right]^2}{\Delta z^2} = 1$$

where

$$D_1 = \min\left[D^S(i-1jk), D^S(i+1jk)\right],$$
$$D_2 = \min\left[D^S(ij-1\ k), D^S(ij+1\ k)\right],$$
$$D_3 = \min\left[D^S(ij\ k-1), D^S(ij\ k+1)\right]$$

The computations were performed level by level until no un-visited grid point was left. The atomic depth was then obtained as tri-linear interpolation of SDFs at the nearest grid points surrounding the atomic centre. We denoted this strategy as Adepth.

Table 1. Vdw radii used for atoms in Adepth

Atomic type	Å
H (polar)	0.225
H (non-polar)	1.320
C	2.175
N	1.850
O	1.800
S	2.200
P	2.150

The performance of Adepth depends on parameters, such as grid spacing, probe radius and atomic vdw radii, used. The finer grid spacing the more accurate results, we found that grid spacing of 0.5 Å is a good compromise between computational cost and accuracy (Figure 2A and B). Accounting for atomic size effects on atom depths, we scaled the SDF magnitude by the atomic extended radius. In the output of PDB format, the atomic depths and scaled ones are presented in B-factor and occupancy columns, respectively. In the web server, an option allows the user to obtain the output of DX format so that one can display 3D-field data in the iso-surface form using VMD (13). Moreover, the user can also request an extruded slice of a macromolecule with pre-defined thickness as the surface skin of the macromolecule (14).

VALIDATION

As there is no analytical solution to quantification of atomic depths for the 3D structure of a protein or a nucleic acid, a novel method is usually validated by comparing with other available methods. A qualitative comparison between various methods on different macro-molecular architectures is presented in the Supplementary Figure S1: a nucleic acid [1BNA (20)], a fibrous protein collagen [1CGD (21)], a large human coagulation factor VIII [3CDZ (22)] and the 30S ribosome subunit [2WDK (23)]. However, the quantitative assessment of these methods remains elusive. A new approach for evaluating the performance of depth computation is presented. A pseudo-molecule was created as a standard structure that consists of one-atom (C^α) residues, namely, UNK (UNK means of unknown type; atomic co-ordinates of the gold-standard molecule in the PDB format are available in the web server). The 3D structure of the standard molecule is formed of four equally spaced (3 Å) concentric spheres as shown in Figure 1. For the present purpose, how dense the residues packed is not the matter, but how much the space filled up is concerned. Totally, there are 1201 residues (or atoms) in each spherical layer, and the outermost sphere has a radius of 12 Å.

As described in 'Implementation' section, probe radius and atomic radii are factors greatly influencing the results of atomic depth. We have performed calculations on the standard molecule with several values of grid spacing Δr (0.5–2.0 Å) such that $\Delta r = \Delta x = \Delta y = \Delta z$, as well as with different probe radii, 1.4–1.6 Å. Among available web

servers, only DEPTH server (6) could take in the standard molecule and go through the entire calculation process, whereas others such as the DPX servers (24) aborted the run because of failure of the associated surface program, whereas the SADIC (25) and PocketDepth servers (5) did not return any results. It should be noted that the inter-sphere distance is measured according to positions of atomic centres, whereas the depth of an atom depends on the distance between the atomic centre and the probe centre. Consequently, to make a direct comparison of depth computations and inter-sphere distances for the standard molecule, one has to subtract the extended radius of an atom from the depth value of that atom. To be explicit, the atomic depth was reduced by 2.175 plus 1.5 Å from Adepth and by 1.53 plus 1.5 Å from the DEPTH server (the value of 1.53 Å was obtained from the carbon-atom type Amber force field, the same force field used to equilibrate the solvent-bath in the DEPTH server). The results of Adepth and DEPTH are plotted in Figure 2B where the data of DEPTH were calculated using default parameters in the website. Adepth results for the probe radius of 1.4 and 1.6 Å are not shown, as their values are close to that of 1.5 Å. Each data point in the plot represents an averaged depth over all atoms in a spherical layer. Despite deviations from true depths become evident at large grid spacing, the order of depth ranking from the Adepth results remains unchanged (Figure 2B).

The computational accuracy of atomic depth is defined as a ratio of measured depth to the true depth. The measured depth was obtained as data values of averaged depth shown in Figure 2B. The dependence of computational accuracy on grid spacing is presented in Table 2. Comparing the depth accuracy, we found that globally Adepth is more accurate than the DEPTH server at small to medium grid spacing (~1 Å). Nevertheless, for the deepest layer, Adepth can obtain better results that DEPTH at fine grid spacing, i.e. 0.1 Å, which demands high CPU cost and huge memory load.

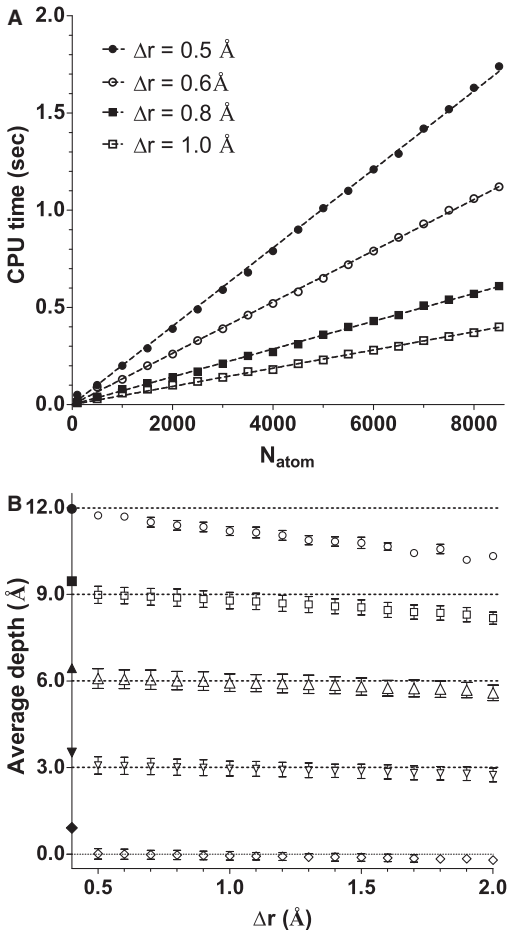

Figure 2. (**A**) Relationship of computational cost and molecular size. The CPU time was plotted against the number of atoms (N_{atom}) of the standard molecule for Δr = 0.5, 0.6, 0.8 and 1.0 Å. The four sets of data values were fitted with linear regression, resulting in a slope of 0.2, 0.1, 0.07 and 0.05 msec/atom for Δr = 0.5, 0.6, 0.8 and 1.0 Å, respectively. (**B**) Validation of Adepth and DEPTH on atomic depth calculations for the standard molecule using a probe radius of 1.5 Å. In the plot, the horizontal dashed lines represent the distances of spherical layers from the surface. Data points with open symbols represent the values of averaged atomic depth versus Δr from Adepth, whereas that with filled symbols located along the vertical axis are referred to the DEPTH server. Bars around the open symbols represent the standard deviation; when it is <1%, bars are not drawn.

DISCUSSION

The relationship between protein interior and atomic depth has been extensively discussed (2,7,25). New implications of atomic depths calculated by the Adepth server will be illustrated later in the text.

Once atomic depths for a macromolecule are calculated, a built-in option in the Adepth server allows the user to obtain a molecular slice based on values of atomic depths. The corresponding outputs are called 'extruded surfaces' that are useful in molecular docking and assembly. It is usually done through an exhaustive computation of inter-atomic distance for detecting atomic overlaps in a complex. In practice, this is not necessary because the desired information is to find bumping atoms at the interface (14). Guided by an extruded surface of a molecular complex, one may reduce the number of atoms in distance calculations by ~4–10 times while maintaining an accurate contour for the complex molecule (Figure 3A).

Another novel feature of Adepth allows the user to visualize 3D iso-surface mapping of atomic depths. As an SDF value carries either sign, it is now possible to accurately display an expanded volume of a given macromolecule.

Table 2. Comparison of computational accuracy between DEPTH and Adepth[a]

Layer	True depth	DEPTH (%)	Δr = 0.5 (%)	Δr = 1.0 (%)	Δr = 2.0 (%)
1	3	17.3	2.6	1.1	8.5
2	6	7.4	1.5	1.0	6.8
3	9	5.1	0.2	2.4	9.1
4	12	0.3	2.2	6.7	13.9

[a]The innermost sphere of the standard molecule is labelled 4 (red in Figure 1), and the layer enumeration is decreased by 1 for next outwards sphere; computations with different grid spacing were performed only by Adepth; true depth and Δr are in the unit of Å.

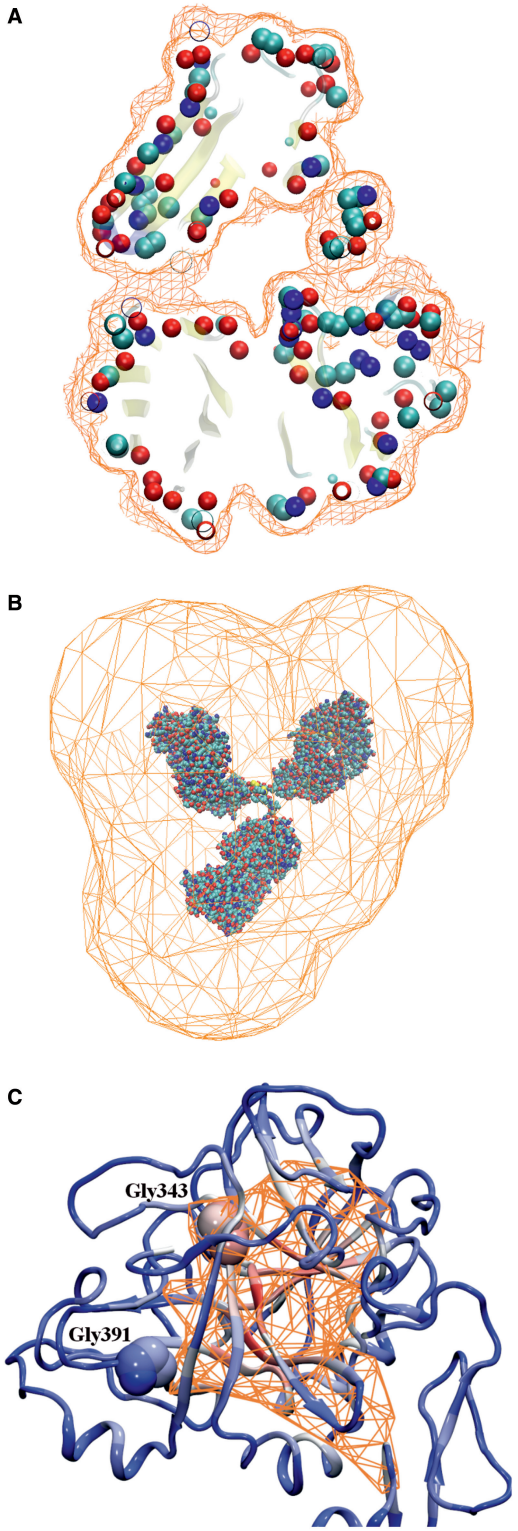

Figure 3. Implications of new representation for atomic depths. (**A**) Cartoon representation of the extruded crystal structure of a Fab domain of antibody protein using a probe radius of 1.5 Å, Δr = 1 Å, and depth selection threshold <3.5 Å. Of 3291 atoms, 805 are found within the given threshold. (**B**) Iso-surface representation for an expanded surface of an antibody molecule. The crystal structure of the antibody is represented by 1IGT (26) with vdw spheres and coloured atomic types: cyan for carbon, red for oxygen, blue for nitrogen and yellow for sulphur atoms. The atomic depths of 1IGT and the associated SDF values at grid points were computed using a probe radius of 1.5 Å at Δr = 1 Å. The iso-surface in orange wireframe was located 50 Å away from the antibody surface. (**C**) Clipped core structure of the serine protease domain of FVIIa (27). The orange wireframe represents the iso-surface of the SDF value at a threshold of 7.5 Å using VMD. The main-chain conformation is displayed in a ribbon of red-to-white-to-blue spectrum, where red and blue are for buried and solvent-exposed regions, respectively. The deepest atom in FVIIa is 13.5 Å inwards from the protein surface. The mutated glycines with almost no solvent-accessible surface areas are represented by coloured CPK spheres. The residue depth was calculated as 10.0 Å for Gly343 and 4.8 Å for Gly391; thus, only Gly343 is enclosed in the orange wireframe.

This is convenient for examining nearby surfaces of a macromolecule while still tracking the distance from the surface atoms. Analogous results could be obtained by traditional computations of ASA; yet, it is not plausible for most surface programs because of the need of a huge probe for generating such a remote surface at 50 Å away as shown in Figure 3B. This feature is useful in AFM for visualizing putative tip-convolution effects. As known, the tip effects lead to dilations of object boundary in AFM images (15,28). With a large radius of tip apex, the tip convolution effect greatly smoothed the macromolecular shape. When an expanded surface of a macromolecule can be displayed, it is easier to grasp an idea of what to expect from AFM imaging on a single molecule (Figure 3B). As tip convolution effect mostly occurred as a 2D artefact in an AFM image, one may scrutinize the strict 3D representation provided by Adepth in certain direction for viewing expected convoluted surfaces for the macromolecule.

The final feature of Adepth is the use of 3D iso-potential surface for constructing a graphic representation of protein cores. Atomic depth is particularly well adapted to the identification of protein core residues as illustrated in the ProCoCoa method (25). As previously suggested, the residue depth is an important parameter for interpreting mutational data in proteins (7) and correlates well with residue hydrophobicity (2,3). Iso-surfaces provide an intuitive interpretation of atomic depth. For instance, graphical representation of atomic depth was helpful in illustrating the puzzling clinical data regarding the impact of a single-nucleotide polymorphism on activities of activated coagulation factor VII (FVIIa). Two apparently similar missense mutations were found in unrelated patients having a gene defect in FVII: Gly343Ser and Gly391Ser; however, the phenotype of Gly343Ser was severe, whereas that of Gly391Ser was minor (29). Using the iso-surface representation of the ternary complex model (27), it can be observed that although both Gly343 and Gly391 are completely buried (Figure 3C), Gly343 is located deeper in the hydrophobic core than Gly391. Thus, the apparent paradox could be interpreted in term of depth rather than solvent-burial. Although this observation remains phenomenological, it conveys the already known idea that a deep mutation as a greater impact on protein structure than a solvent-exposed mutation (7).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figure 1.

ACKNOWLEDGEMENTS

The authors thank Muriel Giansily-Blaizot (Hopital Saint-Eloi, Montpellier) for useful discussions on FVIIa, Michael Odorico, Jean-Marie Teulon and Pierre Parot (CEA-SBTN Marcoule) for AFM imaging and Arnaud Martel (CEA-GIPSI Saclay) for the website developments. The atomic co-ordinates of gold-standard molecule can be downloaded from the server website.

FUNDING

Commissariat à l'Energie Atomique et aux Energies Alternatives; French ANR agency [ANR-07-PCVI-0002-01 to J.-L.P.]. Funding for open access charge: Institutional funds.

Conflict of interest statement. None declared.

REFERENCES

1. Pedersen,T.G., Sigurskjold,B.W., Andersen,K.V., Kjaer,M., Poulsen,F.M., Dobson,C.M. and Redfield,C. (1991) A nuclear magnetic resonance study of the hydrogen-exchange behaviour of lysozyme in crystals and solution. *J. Mol. Biol.*, **218**, 413–426.
2. Pintar,A., Carugo,O. and Pongor,S. (2003) Atom depth as a descriptor of the protein interior. *Biophys. J.*, **84**, 2553–2561.
3. Zhou,H. and Zhou,Y. (2004) Quantifying the effect of burial of amino acid residues on protein stability. *Proteins*, **54**, 315–322.
4. Liu,S., Zhang,C., Liang,S. and Zhou,Y. (2007) Fold recognition by concurrent use of solvent accessibility and residue depth. *Proteins*, **68**, 636–645.
5. Kalidas,Y. and Chandra,N. (2008) PocketDepth: a new depth based algorithm for identification of ligand binding sites in proteins. *J. Struct. Biol.*, **161**, 31–42.
6. Tan,K.P., Varadarajan,R. and Madhusudhan,M.S. (2011) DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins. *Nucleic Acids Res.*, **39**, W242–W248.
7. Chakravarty,S. and Varadarajan,R. (1999) Residue depth: a novel parameter for the analysis of protein structure and stability. *Structure*, **7**, 723–732.
8. Yuan,Z. and Wang,Z.X. (2008) Quantifying the relationship of protein burying depth and sequence. *Proteins*, **70**, 509–516.
9. Song,J., Tan,H., Mahmood,K., Law,R.H., Buckle,A.M., Webb,G.I., Akutsu,T. and Whisstock,J.C. (2009) Prodepth: predict residue depth by support vector regression approach from protein sequences only. *PLoS One*, **4**, e7072.
10. Greer,J. and Bush,B.L. (1978) Macromolecular shape and surface maps by solvent exclusion. *Proc. Natl Acad. Sci. USA*, **75**, 303–307.
11. Hubbard,S.J. and Thornton,J.M. (1993) NACCESS version 2.1.1. Department of Biochemistry and Molecular Biology University College London.
12. Varrazzo,D., Bernini,A., Spiga,O., Ciutti,A., Chiellini,S., Venditti,V., Bracci,L. and Niccolai,N. (2005) Three-dimensional computation of atom depth in complex molecular structures. *Bioinformatics*, **21**, 2856–2860.
13. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
14. Trinh,M.H., Odorico,M., Pique,M.E., Teulon,J.M., Roberts,V.A., Ten Eyck,L.F., Getzoff,E.D., Parot,P., Chen,S.w.W. and Pellequer,J.L. (2012) Computational reconstruction of multidomain proteins using atomic force microscopy data. *Structure*, **20**, 113–120.
15. Trinh,M.H., Odorico,M., Bellanger,L., Jacquemond,M., Parot,P. and Pellequer,J.L. (2011) Tobacco mosaic virus as an AFM tip calibrator. *J. Mol. Recognit.*, **24**, 503–510.
16. Chen,S.w.W., Van Regenmortel,M.H.V. and Pellequer,J.L. (2009) Structure-activity relationships in peptide-antibody complexes: implications for epitope prediction and development of synthetic peptide vaccines. *Curr. Med. Chem.*, **16**, 953–964.
17. MacKerell,A.D., Bashford,D., Bellott,M., Dunbrack,R.L. Jr, Evanseck,J.D., Field,M.J., Fisher,S., Gao,J., Guo,H., Ha,S. *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
18. Osher,S. and Fedkiw,R. (2003) *Level Set Methods and Dynamic Implicit Surfaces*. Springer-Verlag, New York.
19. Tsitsiklis,J.N. (1995) Efficient Algorithms for Globally Optimal Trajectories. *IEEE Trans. Auto Control*, **40**, 1528–1538.
20. Drew,H.R., Wing,R.M., Takano,T., Broka,C., Tanaka,S., Itakura,K. and Dickerson,R.E. (1981) Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl Acad. Sci. USA*, **78**, 2179–2183.
21. Bella,J., Brodsky,B. and Berman,H.M. (1995) Hydration structure of a collagen peptide. *Structure*, **3**, 893–906.
22. Ngo,J.C., Huang,M., Roth,D.A., Furie,B.C. and Furie,B. (2008) Crystal structure of human factor VIII: implications for the formation of the factor IXa-factor VIIIa complex. *Structure*, **16**, 597–606.
23. Voorhees,R.M., Weixlbaumer,A., Loakes,D., Kelley,A.C. and Ramakrishnan,V. (2009) Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat. Struct. Mol. Biol.*, **16**, 528–533.
24. Pintar,A., Carugo,O. and Pongor,S. (2003) DPX: for the analysis of the protein core. *Bioinformatics*, **19**, 313–314.
25. Bottini,S., Bernini,A., De Chiara,M., Garlaschelli,D., Spiga,O., Dioguardi,M., Vannuccini,E., Tramontano,A. and Niccolai,N. (2013) ProCoCoA: A quantitative approach for analyzing protein core composition. *Comput. Biol. Chem.*, **43**, 29–34.
26. Harris,L.J., Larson,S.B., Hasel,K.W. and McPherson,A. (1997) Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry*, **36**, 1581–1597.
27. Chen,S.w.W., Pellequer,J.L., Schved,J.F. and Giansily-Blaizot,M. (2002) Model of a ternary complex between activated factor VII, tissue factor and factor IX. *Thromb. Haemost.*, **88**, 74–82.
28. Tian,F., Qian,X. and Villarrubia,J.S. (2008) Blind estimation of general tip shape in AFM imaging. *Ultramicroscopy*, **109**, 44–53.
29. Pinotti,M., Etro,D., Bindini,D., Papa,M.L., Rodorigo,G., Rocino,A., Mariani,G., Ciavarella,N. and Bernardi,F. (2002) Residual factor VII activity and different hemorrhagic phenotypes in CRM(+) factor VII deficiencies (Gly331Ser and Gly283Ser). *Blood*, **99**, 1495–1497.